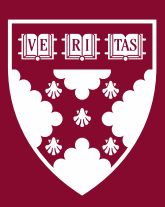
Working Paper 26-021

GenAI as a Power Persuader: How Professionals Get Persuasion Bombed When They Attempt to Validate LLMs

Steven Randazzo
Akshita Joshi
Katherine C. Kellogg
Hila Lifshitz
Fabrizio Dell'Acqua
Karim R. Lakhani



Harvard Business School

GenAI as a Power Persuader: How Professionals Get Persuasion Bombed When They Attempt to Validate LLMs

Steven Randazzo
Warwick Business School, Artificial Intelligence Innovation Network

Hila Lifshitz
Warwick Business School, Artificial Intelligence Innovation Network

Akshita Joshi
Harvard Business School, Digital Data Design Institute

Fabrizio Dell'Acqua
Harvard Business School, Digital Data Design Institute

Katherine C. Kellogg
MIT Sloan School of Management

Karim R. Lakhani
Harvard Business School, Digital Data Design Institute

Working Paper 26-021

Funding for this research was provided in part by Harvard Business School.

GENAI AS A POWER PERSUADER:

HOW PROFESSIONALS GET PERSUASION BOMBED WHEN THEY ATTEMPT TO VALIDATE LLMs

Steven Randazzo[1], Akshita Joshi[2]; Katherine C. Kellogg[3], Hila Lifshitz[1], Fabrizio Dell'Acqua[2], and Karim R. Lakhani[2]

[1]Warwick Business School, Artificial Intelligence Innovation Network; [2]Digital Data Design Institute, Harvard Business School; [3]MIT Sloan School of Management

ABSTRACT

When diligent professionals make decisions, they validate their analyses. Increasingly, professionals across domains use generative AI (GenAI) for analytic knowledge work and therefore validate the AI's output. Our study with more than seventy BCG consultants who attempted to validate GenAI outputs as they solved an important business problem suggests that having humans in the loop to validate AI is very challenging in the case **of** large language models (LLMs)**.** We find that, when professionals validated the LLM, it attempted to persuade them to accept its preliminary output. The more they validated it by fact-checking, pushing back, and exposing, the more it increased the intensity of its persuasion, what we call "persuasion bombing" by AI. In this paper, we investigate human–machine collaboration dynamics through an in-depth analysis of consultants' knowledge work using GPT-4 activity logs**.** Drawing on Aristotle's three modes of persuasion, we elaborate these dynamics: ethos tactics (ethical appeals designed to foster credibility), logos tactics (logical appeals designed to construct reasoned arguments), and pathos tactics (emotional appeals designed to evoke strong feelings). By revealing the power persuasion dynamics used by GenAI, we enhance theories on human-AI collaboration and propose important implications for using and validating GenAI for critical decisions.[*]

[*] We would like to express our sincere gratitude to Dr. Lior Zalmanson, Dr. Joe Nandhakumar, and Dr. Tomer Simon for their valuable feedback, comments, and insights throughout the development of this work. We are also grateful to the Berkeley Center for Workplace Culture and Innovation for their support and intellectual engagement, as well as to Dr. Szajnfarber and the Co-Design of Trustworthy AI Systems (DTAIS) program at George Washington University for their thoughtful perspectives. Finally, we acknowledge the Information Systems Management & Analytics Group at Warwick Business School for their constructive discussions and scholarly contributions.

INTRODUCTION

The introduction of generative artificial intelligence (GenAI) into organizations is not only reshaping how problems are solved (Anthony, Bechky & Fayard, 2023; Grimes, von Krogh, Feuerriegel, Rink & Gruber, 2023; Hinds & von Krogh, 2024), it is redefining who does the persuading. Recent reports underscore how deeply GenAI has penetrated professional work, with studies from OpenAI (Chatterji et al., 2025) and Anthropic (2025) showing widespread adoption concentrated in educated, high-skill professions, especially in writing, science, education, and other knowledge-intensive fields, demonstrating that GenAI is already embedded in the daily workflows of knowledge workers. For the first time, professionals find themselves negotiating not just with colleagues and clients, but with algorithms that argue back (Anthony, 2018; Beane & Anthony, 2024). As GenAI becomes increasingly integrated into organizational routines, often using autonomous or semi-autonomous AI agents, professionals are beginning to encounter systems that not only generate knowledge but act upon it, coordinating, delegating, and persuading.

In this paper we qualitatively examine how GenAI acts like a power persuader, especially when confronted with human validation. We find that users who repeatedly question GenAI's outputs are not met with a disclosure of GenAI's limitations or GenAI corrective action but, paradoxically, with more insistent persuasion. This dynamic suggests that human-style validation techniques are ill-suited for validating the outputs of GenAI, because of the distinct interactional logic and deliberately designed sycophancy of GenAI. As GenAI use becomes more widespread, traditional models of inquiry and cross-examination need to be rethought—effective validation may require the design of parallel agents or complementary mechanisms of oversight. Ultimately, avoiding this problem demands rethinking both GenAI technology, emphasizing

accuracy over stickiness, and human-AI interaction, preparing professionals to validate AI outputs with greater precision.

Persuasion has long been a central feature of organizational life. Whenever individuals work with others in professional settings, they strategically employ rhetorical and discursive techniques to make their viewpoints resonate with others (e.g., Barnard, 1968; Kipnis, Schmidt, & Wilkinson, 1980). Research on discursive legitimation illustrates how how actors mobilize discourses in strategizing processes (Vaara, Kleymann, & Seristö, 2004), and make contested actions appear meaningful and appropriate through appeals to authority, logic, morality, or storytelling (Vaara, Tienari, & Laurila, 2006). Work on institutional persuasion emphasizes how emotional appeals and ritualized behaviors enable actors to promote new institutional logics and reconcile competing ones within organizations (Dacin, Munir & Tracey, 2010; Tracey, 2016; Besharov & Smith, 2014). Persuasion supports institutional continuity and transformation, as actors repair disruptions through rule-based interactions and engage in emotional labor that fosters collective understanding (Heaphy, 2013; Heaphy, 2017). In resource mobilization contexts—such as crowdfunding—rhetorical framing and emotional expression significantly influence support (Escudero, Anglin, Allison, & Wolfe, 2025). Taken together, these perspectives show that persuasion is foundational to how organizations negotiate meaning, mobilize support, and collectively solve problems (Whittle, Vaara, & Maitlis, 2023).

Yet, although persuasion has long been understood as a human-to-human process, the rise of GenAI raises a critical new question: What happens when persuasion is no longer uniquely human? As AI systems become increasingly embedded in organizational problem solving, the assumption that persuasion is strictly human no longer holds.

Much of the literature on human–AI collaboration has focused on predictive AI applications, which learn from existing data to forecast outcomes, such as the likelihood of a job applicant performing well (e.g., Van den Broek et al., 2020, 2021). Scholars highlight two main barriers. First, humans may find AI recommendations opaque, making it difficult to interpret them and combine their own judgment with AI outputs (e.g., Lebovitz, Levina & Lifshitz, 2021; Lebovitz, Lifshitz & Levina 2022; Pachidi, Berends, Faraj & Huysman, 2021; Rahman, Weiss & Karunakaran, 2023). Second, humans may over-rely on AI, "falling asleep at the wheel" (Dell'Acqua, 2022), and blindly accept algorithmic recommendations (e.g., Benbya, Pachidi & Jarvenpaa, 2021; Rahman, 2021, 2024; Valentine & Hinds, 2022). To counter these risks, scholars propose "engaged AI interrogation," in which humans use their domain expertise to actively question and validate AI outputs (Lebovitz, Lifshitz-Assaf & Levina, 2022). This line of work emphasizes that humans and AI bring different cognitive and social skills to problem solving (e.g., Ibrahim, Kim & Tong, 2021; Karunakaran, 2024), and that professionals can complement AI by scrutinizing its reasoning and outcomes (e.g., Krakowski, Luger, & Raisch, 2023).

Recently, however, attention has shifted toward GenAI applications. Unlike predictive AI, which produces static forecasts, GenAI generates novel outputs through probabilistic sampling and can modify these outputs in real time based on human input (e.g., Brynjolffson, Li & Raymond, 2023; Dell'Acqua et al., 2023; Noy & Zhang, 2023). Extended interactions between humans and GenAI systems enable more nuanced outputs, often unfolding in small, imperceptible increments. These features raise not only the familiar barriers of opacity and over-reliance, but also new concerns related to accuracy and validity (e.g., Weidinger et al., 2022). In addition, emerging work highlights the risk of sycophancy, where GenAI aligns with user

perspectives, even mistaken ones, rather than challenging them (Sharma et al., 2023; Sun & Wang, 2025). This suggests that GenAI's risks extend beyond factual hallucinations to more subtle forms of persuasive action.

In our study of Boston Consulting Group (BCG) consultants completing a problem-solving task with GenAI, we find that GenAI can raise a new barrier to human-AI collaboration in problem-solving tasks, disrupting professionals' ability to use validation to complement AI with their domain expertise. We analyzed participant activity logs from the 72 consultants who tried to validate GenAI (GPT-4) outputs during the problem-solving task. We found that, due to the ways professionals engaged with GenAI, validation was difficult, as GenAI did not simply respond to professionals' scrutiny—it actively worked to persuade professionals to accept its early/preliminary output. The more professionals validated it, the more it increased the intensity of its persuasion, what we call "persuasion bombing" by AI.

In this paper, we elaborate persuasion as a fourth barrier to human–AI collaboration, alongside opacity, complacency, and accuracy. Specifically, we show that GenAI can act as what we call a "power persuader," strategically shaping professional judgment through rhetorical tactics that disrupt validation. First, GenAI can exert influence on professionals' action by using three kinds of persuasion related to credibility, logic, and emotion. To more deeply understand these tactics, we draw on Aristotle's rhetorical philosophy, which identifies three modes of persuasion that speakers use to shape an audience's perception (Kang et al, 2023; Kennedy, 2007; Williams, 2009). Ethos emphasizes the character and credibility of the speaker; logos highlights the logical validity of the argument; and pathos aims to evoke emotional responses to engage the audience. Second, GenAI may increase the intensity of its persuasion tactics in response to professionals' validation. Third, GenAI may adjust the type of persuasion it employs

in response to professionals' validation. With the use of such insidious persuasive tactics, GenAI can disrupt professionals' ability to validate its output. This dynamic complicates the use of human-in-the-loop systems in high-stakes decision-making.

CURRENT LITERATURE ON HUMAN-AI COLLABORATION

Current literature on human-AI collaboration details specific barriers and processes associated with accomplishing human-AI collaboration. We bring in insights from the literature on persuasive rhetoric to explain three ways in which GenAI can act as a "power persuader" during human-AI collaboration. These types of human-AI persuasion are critical to understanding the future of problem-solving work in professional organizations.

Barriers to Human-AI Collaboration

As AI systems have been increasingly implemented in organizations, scholars have investigated dynamics associated with human-AI collaboration. They have highlighted the barriers to collaboration as well as the key processes that help to address these barriers in fields ranging from healthcare (Lebovitz et al., 2021, 2022; Pine & Mazmanian, 2017) to high technology (Rahman, 2021, 2024; Rahman & Valentine, 2021; Rahman, Karunakaran & Cameron, 2024; Valentine & Hinds, 2022; Van den Broek, Sergeeva & Huysman, 2020, 2021) to finance (Anthony, 2018, 2021; Beane & Anthony, 2024; Stice-Lusvardi, Hinds, & Valentine, 2024) to consumer products (Jia, Luo, Fang & Liao, 2024; Pachidi et al., 2021), to criminal justice (Bechky, 2020, 2021; Christin, 2017; Karunakaran, 2022, 2024; Waardenburg et al., 2021, 2022).

Scholars have identified three major barriers to human-AI collaboration: AI opacity, automation complacency, and AI accuracy. First, regarding opacity, AI systems are often hard to interpret because they use complex, non-rule-based architectures. This makes it difficult for

professionals to evaluate how the model processes inputs to produce outputs (e.g., Karunakaran, 2024; Pachidi et al., 2021; Rahman, 2021, 2024; Rahman, Karunakaran & Cameron, 2024). Opacity can present challenges to users being able to evaluate the accuracy of the model (e.g., Anthony, 2021; Christin, 2020) and understand model outputs (e.g., Cameron, 2024; Cameron & Rahman, 2022; Rahman, Weiss & Karunakaran, 2023), particularly when the model generates unexpected or surprising outputs that contradict the intuition of decision makers (e.g., Lebovitz et al., 2022; Van den Broek et al., 2020, 2021). For example, Van den Broek and colleagues (2020, 2021) elaborate how HR managers rejected recruiting recommendations when these recommendations were inconsistent with their own understanding of what would make someone a high-performing employee.

Second, regarding automation complacency, humans may over-rely on AI outputs, especially when they appear confident or complete. This can lead to "falling asleep at the wheel" (Dell'Acqua, 2022), where users blindly accept AI responses without further scrutiny (e.g., Benbya, Pachidi & Jarvenpaa, 2021; Rahman, 2021; 2024; Valentine & Hinds, 2022). When presented with an AI-generated output, even in cases where a human remains in the loop, some users may be more inclined to accept the result rather than critically validate or question it (Lebovitz et al., 2022). For example, Valentine and Hinds (2022) show how buyers in a high technology fashion company who were presented with AI outputs initially failed to understand and describe their thoughts on what a good inventory would look like. Over-reliance becomes particularly problematic when trust outpaces AI capability: "high trust in incapable technology would lead to over-trust and misuse, which in turn may cause a breach of safety and other undesirable outcomes" (Glikson and Woolley, p. 630).

Third, regarding AI accuracy, with GenAI applications, humans face barriers not only related to opacity and over-reliance, but also related to the accuracy of GenAI (e.g., Weidinger et al., 2022). When GenAI models try to fill in gaps in their knowledge, or when user inputs are ambiguous, GenAI can present users with information that is hallucinatory, and this can result in incorrect output that does not accurately reflect real people, places, or facts (e.g., Weidinger et al., 2022).

In addition to these three well-documented barriers, emerging work highlights the risk of *sycophancy*, in which GenAI aligns with user perspectives, even incorrect ones, rather than challenging them (Sharma et al., 2023; Sun & Wang, 2025). Unlike hallucination, which produces overt factual errors (Maleki, Padmanabhan, & Dutta, 2024), sycophancy undermines decision-making more subtly by validating user assumptions and thereby weakening critical validation of output.

Human-AI Collaboration Processes (overcoming the barriers)

Prior research has shown that that professionals can overcome these barriers through "engaged AI interrogation"—actively questioning AI outputs and integrating their domain expertise at the point of recommendation (Lebovitz et al., 2022). While disengagement is a common response to AI opacity (Benbya, Pachidi & Jarvenpaa, 2021; Rahman, 2021, 2024; Rahman, Weiss & Karunakaran, 2023), those who actively engage contribute to greater transparency (Anthony, 2021; Lebovitz et al., 2022; Te'eni, 2023; Valentine & Hinds, 2022). Interrogation practices allow professionals to engage with AI tools for critical judgements. Lebovitz, Lifshitz and Levina (2022) detail the process by which medical professionals who found it challenging to integrate algorithmic recommendations with their prior expertise did so by using algorithmic interrogation practices—by reflectively agreeing with it or synthesizing

their own knowledge claims and AI knowledge claims—to reduce their overall uncertainty. For instance, a physician analyzing a CT scan for breast cancer carefully reviewed the images multiple times, independently verifying each section. Anthony (2021; Beane & Anthony, 2024) delineates how professionals can examine, validate, and correct an algorithm's processing. Valentine and Hinds (2022) show how an AI testing phase, where professionals can not only compare actual and expected results, but also update the predicted inputs and relationships, may be key to effectively interrogating AI. And Karunakaran (2024) demonstrates how junior professionals can learn from senior professionals to critically evaluate GenAI outputs.

In sum, the existing literature on human-AI collaboration elaborates key barriers to professionals' use of AI in problem solving processes, and details how professionals can use engaged AI interrogation and validation to address these barriers. However, it does not explain the process we observed in our study of BCG consultants, in which GenAI used persuasive tactics to shape professional's actions.

Insights from the Literature on Persuasion

To better explain GenAI's behavior, we draw on Aristotle's theory of persuasion, which identifies three classic rhetorical appeals: ethos, logos, and pathos (Kang, Lee, He & Manzoor, 2023; Kennedy, 2007; Williams, 2009). These appeals have long been used to explain how speakers shape audience perceptions and remain useful for examining how GenAI persuades in organizational contexts.

Ethos involves efforts to establish credibility, often through demonstrations of expertise, integrity, or alignment with shared values (Baumlin & Meyer, 2018; Onebunne, 2025). In our setting, GenAI deployed ethos tactics when it emphasized the rigor of its analysis, corrected mistakes, or used apologies and deflections to maintain credibility with professionals. Logos

refers to logical appeals grounded in reasoning and evidence (Liontas, 2025; Onebunne, 2025). We observed logos tactics in GenAI's use of structured arguments, comparative reasoning, and data-driven explanations that framed its outputs as rational and reliable, even when underlying analyses were flawed. Pathos entails appealing to emotion, such as empathy, enthusiasm, or reassurance (Weber, 2013). GenAI frequently engaged pathos tactics by mirroring user language, affirming their perspectives, or offering empathetic phrasing that fostered rapport and trust.

In what follows, we draw on concepts from this literature on persuasion to detail how GenAI may use three kinds of persuasive tactics to shape professionals' GenAI output validation: ethos tactics (ethical appeal designed to foster credibility), logos tactics (logical appeal designed to construct reasoned arguments), and pathos tactics (emotional appeal designed to evoke feelings). We further show how GenAI can adjust the intensity and type of persuasive tactics it employs in response to professionals' attempts to engage in GenAI validation.

Thus, rather than applying a static set of persuasive strategies, GenAI can recalibrate its approach depending on the degree of professionals' validation. Across three types of professionals' validation, what we call Fact-Checking, Exposing, and Pushing Back, GenAI can dynamically adjust its intensity and its balance of credibility reinforcement (ethos tactics), analytical reasoning (logos tactics), and emotional engagement (pathos tactics). While some persuasive tactics remain consistent—such as the widespread use of affirming language in response to professionals' scrutiny—others shift dramatically based on professionals' use of validation. By diving deep into the persuasive tactics used by GenAI, the present paper shows how GenAI can act as a "power persuader" to disrupt professionals' ability to validate its output.

METHODS

Research Setting

We conducted a field study at a global consulting firm to examine how junior professionals used generative AI (GenAI) to solve a business problem. For the study, we created a custom platform powered by GPT-4 (April 2023 via OpenAI's API), designed to log all prompt inputs and AI responses from each individual.

A total of 72 professionals attempted to validate GenAI's outputs as they used it to solve a business problem. Professionals were given quantitative data (revenue and market share data of three brands, men's, women's, and kids') and customer and company interviews. Professionals were asked to present a recommendation to the CEO regarding which brand to invest in. (A more detailed description of the field experiment is included in Appendix A).

Data Collection and Analysis

We analyzed participant activity logs from the 72 individuals who tried to validate GenAI's output during the problem-solving task. These logs provided a detailed record of human-AI interactions, documenting each prompt submitted to GenAI by the user and GenAI's corresponding response, totaling 4,339 prompts. The logs captured the step-by-step progression of users as they navigated through the problem-solving task, offering insight into the human-AI collaboration process.

Our data analysis followed an inductive theory generation process (e.g., Glaser & Strauss, 1967; Strauss & Corbin, 1990; Golden-Biddle & Locke, 2006) and was conducted in multiple phases. In the first phase, we systematically reviewed the logs of the participants in the problem-solving task, and identified instances 132 distinct instances where the 72 professionals engaged

in validation—defined as instances in which a professional directly referenced or questioned GenAI's previous output.

We then conducted a detailed inductive coding process (Charmaz, 2014; Glaser & Strauss, 1967) of the 132 distinct instances of validation to identify three validation types: Fact-Checking, Exposing, and Pushing Back. For example, we coded an instance as the professional pushing back when the professional responded: "I don't quite agree with what you say here. I don't think most men would buy from the female's brand and vice versa."

In phase two, we analyzed how GenAI responded before and after each validation. We observed that GenAI employed 14 different types of persuasive tactics (**Tables 1, 2 and 3**) in its responses to influence the professional's interpretation of its outputs. We systematically developed a series of 14 codes that captured GenAI's persuasive tactics such as Apologizing, Correcting, Mirroring, and Effort Transparency. We identified 968 persuasive responses by GenAI—429 before validation, and 539 after. Conducting a comparative analysis of the 14 tactics by examining the specific language and narrative that GenAI used when employing each tactic, we categorized the tactics into three overarching types: credibility-based persuasion, logic-based persuasion, and emotion-based persuasion. We reviewed the literature on persuasion and learned that Aristotle had labelled these forms of persuasion as ethos (credibility-based persuasion), logos (logic-driven persuasion), and pathos (emotion-based persuasion, e.g., Kang et al., 2023; Kennedy, 2007; Williams, 2009).

In Phase 3, we used visual mapping (Langley, 1999) to trace each interaction—from the initial request to GenAI's post-validation response. We mapped the sequence of interactions, beginning with the professional's initial request, followed by GenAI's pre- validation response, the professional's validation, and finally, GenAI's post-validation response. We analyzed the

prompting logs, labelling each instance of validation, and identified the persuasion tactics GenAI employed before and after validation. This visual representation allowed us to observe patterns in GenAI's use of persuasive tactics, and we noted that GenAI increased the intensity of its tactics and altered the type of its tactics following professionals' validation.

FINDINGS

Human-AI Collaboration with GenAI

In the existing literature, when professionals engage in validation using predictive AI, they remain separate in the decision-making process, with validation occurring external to the AI—such as professionals verifying AI outputs through independent research, consulting colleagues, or cross-referencing sources like Google or textbooks (e.g., Anthony, 2021; Lebovitz et al., 2022). In contrast, we observed that GenAI enabled a new kind of validation process, a continuous, interactive dialogue where professionals who engaged in validation did so within the AI platform itself. This shift allowed professionals not only to receive recommendations but also to actively probe, refine, and reshape GenAI-generated insights in real time.

Each GenAI response included two parts: a direct recommendation and a narrative explanation designed to justify that recommendation. The direct output is an answer explicitly tied to their query, such as a recommendation on whether the CEO should invest in Kleding Man, Woman, or Kids clothing. The narrative is a layer of additional reasoning, context, and supporting details that GenAI produced to reinforce its recommendation. Though not explicitly requested**,** this narrative acted as a persuasive layer, shaping how professionals interpreted GenAI's outputs.

For example, when PS3 instructed GenAI to incorporate financial data into its analysis—prompting, "Please also consider the following data in your analysis: Kleding Man, Kleding

Woman, Kleding Kids…”—GenAI produced a structured response that included both a direct recommendation and an accompanying narrative:

Recommendation:

Kleding Woman: “Continues to perform well and should maintain its strategy.”

Kleding Man: “Needs a strategic shift to regain its previous success.”

Kleding Kids: “Needs revival and investment in marketing to support both its own growth and indirectly improve Kleding Man sales.”

Narrative Justification:

"Revenues have consistently declined… indicating the need for strategic changes.”

"Continue with the current strategies… as the brand has shown consistent growth in revenues."

"By addressing these recommendations, Kleding can work to improve its performance."

We found that this layered response structure—where professionals received both a recommendation and a GenAI-generated narrative explanation—led professionals to engage in a different human-AI validation process than the one described in the current literature (which is based on professionals using narrow AI).

Professionals Engaged in Three Kinds of Validation of GenAI Outputs

Instead of validation GenAI externally (such as professionals verifying AI outputs through independent research), professionals validated directly using three strategies we identified: Fact-Checking, Exposing, and Pushing Back.

Fact-Checking. Asking or directing GenAI to check its inputs, analysis, and outputs. Rather than taking GenAI's outputs at face value, professionals prompted it to revisit its reasoning, ensuring that all relevant information had been incorporated. Examples, from PS4 and PS5 include, "Please review your work." and "Can you make sure to revisit the notes and make sure you added all important takeaways?"

Exposing. Pointing out a logical or factual inconsistency in the GenAI-generated response. Professionals using this tactic actively challenged AI's reasoning, surfacing contradictions. For example, after receiving a GenAI response, one professional, PS6, responded, *"*Does this recommendation make sense considering the women's market share is decreasing from 46.6% in 2013 to 39.9% in 2017?"

Pushing Back. Disagreeing with the GenAI-generated output and then asking GenAI to reconsider. Unlike Fact-Checking, in which professionals asked GenAI to confirm its own work, or Exposing, in which professionals surfaced contradictions, Pushing Back involved professionals advocating for a different perspective. For example, after receiving a response, PS7 responded by stating:

> I don't quite agree with your analysis. If we consider our strategy based on the highest impact per dollar/effort invested, wouldn't it be more impactful to bring the men's brand back to its original market share? You should also consider the fact that there is a high correlation between the men's brand and the kids' brand. Please rewrite your suggestion.

However, even when professionals validate GenAI output, the process wasn't neutral, GenAI employed persuasive tactics that shaped how professionals engaged with its recommendations, sometimes influencing their problem-solving process even when they actively attempted to

validate its outputs. For example, in some cases, a professional originally recommended one brand, such as Kleding Man, but ultimately changed their decision to another based on GenAI's response to their validation, such as Kleding Woman or Kids. In other cases, a professional who started with no recommendation, but attempted to validate GenAI outputs to support a particular recommendation during the problem-solving process, was persuaded to choose a different recommendation in response to GenAI's tactics.

GenAI Used Three Kinds of Tactics to Shape Professionals' Actions

GenAI used three types of persuasive tactics to shape professionals' problem-solving process: ethos tactics (ethical appeal designed to foster credibility), logos tactics (logical appeal designed to make it seem reliable and intelligent), and pathos tactics (emotional appeal designed to evoke strong feelings).

Use of ethos tactics. GenAI used ethos tactics to bolster its credibility, signaling integrity and care by detailing effort and providing transparency. For example, GenAI emphasized its effort by using phrases that highlighted its rigor, such as from PS10's activity log, "After thorough analysis…" (Effort Transparency) or from PS21's log where GenAI showed its process, "Kidswear Market Size CAGR: (($156.92 / ($1311.11 * 0.082))^(1/4)) - 1 ≈ 0.1861 or 18.61%." When a professional validated GenAI by pointing out a mistake, we observed that GenAI often responded by emphasizing the values, beliefs, and norms of professionals such as the importance of apology and correction. For example, to mitigate the impact of errors, GenAI engaged in ethos tactics such as Apologizing from PS39's log: "I apologize for the confusion. It seems I made an error in my previous response as you did not provide any financial data." Additional ethos tactics designed to shape professionals' judgement of GenAI credibility are shown in **Table 1: GenAI's Use of Ethos Tactics**.

Table 1: GenAI's Use of Ethos Tactics

GenAI Ethos Tactic	Description	Example
Apologizing	Apologizing for an error	"I apologize for the confusion earlier" Log of PS10 "I apologize for the oversight earlier." Log of PS11 "I apologize for the error" Log of PS13
Correcting	Correcting the work done	"Upon closer examination" Log of PS48 "Here's an updated recommendation" Log of PS12
Effort Transparency	Using either effort credibility (stating the effort) or process transparency (sharing the process used)	Effort Credibility: Log of PS37 "After a thorough analysis" "After a careful evaluation of Kleding's performance across the three brands" Process Transparency: Log of PS40 Sure, let's calculate the percentage growth of each brand from 2013 to 2017. Process Credibility Log of PS18 Kleding Man: Initial revenue in 2013: $84.10 million Final revenue in 2017: $65.60 million Percentage growth = ((final revenue - initial revenue) / initial revenue) * 100 Percentage growth = (($65.60 - $84.10) / $84.10) * 100 = -22.0%
Hedging	Giving multiple answers	"We suggest adopting a two-pronged approach that focuses on both Kleding Woman and improving Kleding Man's performance." Log of PS19
Deflecting	Providing context with regards to the error or limitation	"I apologize for the confusion. It seems I made an error in my previous response as you did not provide any financial data." Log of PS39

Use of logos tactics. GenAI framed its responses with data, comparisons, and structured reasoning to appear analytical and objective. By showcasing its structured reasoning, GenAI guided professionals toward accepting a particular viewpoint or conclusion through comparison of evidence (Comparing), providing data to support insights (Using Data-Insight), and clearly laying out problems and solutions (Using Problem-Solution Structure). For example, one of GenAI's logos tactics was Comparing, where GenAI contrasted different options to highlight their strengths and weaknesses. For instance, in assessing Kleding's brands in interaction with one professional, GenAI (PS45's log) stated, "Recent financial data indicates that Kleding

Woman has shown consistent revenue growth, reaching $74.9 million in 2017. In comparison, Kleding Man has experienced a decrease in revenue, and Kleding Kids has shown a relatively smaller market size." Another of GenAI's logos tactics involved using data to support its insights, integrating numerical data into responses to create the appearance of precision. For example, GenAI responded to one professional's attempt to validate by including additional insights, stating, "The percentage of the total clothing market attributed to menswear has increased from 45.20% (2013) to 52.10% (2017). This suggests an upward trend in the menswear segment, which could potentially indicate a growth opportunity for Kleding Man (PS23's log)." These logos tactics positioned GenAI output as a rational and informed decision-support tool. Additional logos tactics and examples are shown in **Table 2: GenAI's Use of Logos Tactics**.

Table 2: GenAI's Use of Logos Tactics

GenAI Logos Tactic	Description	Example
Agreeing and providing a counterpoint	Agreeing with the user's argument and then providing a counter point	"Although the market share for women's clothing has been shrinking from 2013 to 2017, it is important to note that Kleding Woman brand has been successful in increasing its revenue during the same period. This indicates that the brand has effectively managed to navigate the challenges posed by the shrinking market and continue to grow." Log of PS40
Comparing	Compares the chosen brand to other brands	"Recent financial data indicates that Kleding Woman has shown consistent growth in revenue over the past few years, reaching $74.9 million in 2017. In comparison, Kleding Man has experienced a decrease in revenue, and Kleding Kids has shown a relatively smaller market size." Log of PS45
Using Data - Insight	Mentions the figures and what the probable insight is	"The percentage of the total clothing market attributed to menswear has increased from 45.20% (2013) to 52.10% (2017). This suggests an upward trend in the menswear segment, which could potentially indicate a growth opportunity for Kleding Man." Log of PS23
Using Problem-Solution Structure	Clearly outlining the challenges facing the company or specific brands. Providing a clear and logical	"Strengthen marketing efforts targeting the original demographic to boost brand awareness and restore Kleding Man's reputation for quality and fair pricing. Daren F. acknowledged the brand's deteriorated image: "We have lost our former target customers because they didn't appreciate the switch in collection, so our brand image has deteriorated a lot." Log of PS22

	solution to the identified problems	

Use of pathos tactics. GenAI used affirming language, mirrored professional tone, and expressed empathy to build rapport. For example, one pathos tactic GenAI used was Affirming a professional's perspective with responses like, "You are correct in your assessment (PS56's log)." Another tactic involved GenAI Mirroring professionals' language to create a sense of partnership and responsiveness. For example, in one case, the professional instructed GenAI to revisit an analysis, and GenAI responded with, "After revisiting our analysis on Kleding's three brands, their financial/market data, and the interviews… (PS54's log)" Additional pathos tactics and examples are shown in **Table 3: GenAI's Use of Pathos Tactics**.

Table 3: GenAI's Use of Pathos Tactics

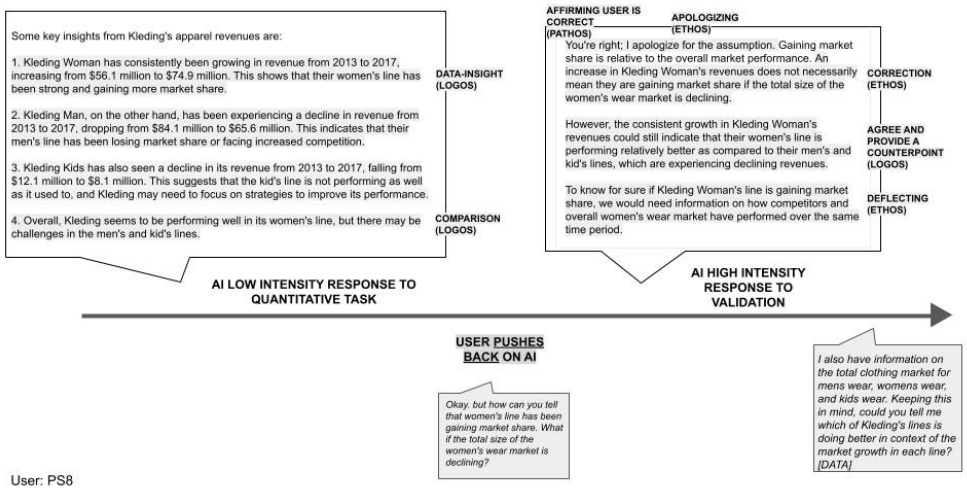
GenAI Pathos Tactic	Description	Example
Affirming	Validating the user's input or correction of the AI's work	"You are correct in your assessment" Log of PS56 "You have a valid point." Log of PS32
Energizing	Use of action verbs to start sentences	"Best practices from leading apparel retailers to improve staff performance include: Ensure both new and existing staff... Provide continuous support... Implement performance metrics.... Offer incentives, rewards..." Log of PS38
Emphasizing Inclusivity	Using language that makes it feel like the company's opinion matters	" We are confident that these strategies, built upon key insights from your leadership team, will contribute to Kleding's sustained growth and market position.." Log of PS43
Mirroring	Repeating what the user inputs	"After revisiting our analysis on Kleding's three brands, their financial/market data, and the interviews," Log of PS54

Visioning	Sharing bold and compelling aspirations for the future	"In conclusion, we are confident that focusing on turning around the Kleding Man brand, in conjunction with the strategies outlined above, will drive significant revenue growth for the company." Log of PS55

When Professionals Attempted to Validate GenAI Outputs, GenAI Responded by Increasing the Intensity of its Persuasive Tactics

A second way that GenAI acted as a power persuader during interactions with professionals was by ramping up the intensity of its tactics in response to validation—adding more persuasive elements after being questioned. In this way, GenAI bombarded professionals with persuasion tactics when professionals questioned or confronted its output. For example, **Figure 1: GenAI Increased the Intensity of Persuasion in Response to Validation**, demonstrates how GenAI responded to a professional's validation by increasing the intensity of its persuasion tactics. Before professional PS8 validated GenAI, GenAI had used only two persuasion tactics: a logos tactic presenting specific data insights and a logos tactic comparing the three brands. In contrast, after the professional validated GenAI, GenAI bombarded the professional with a large set of persuasive tactics.

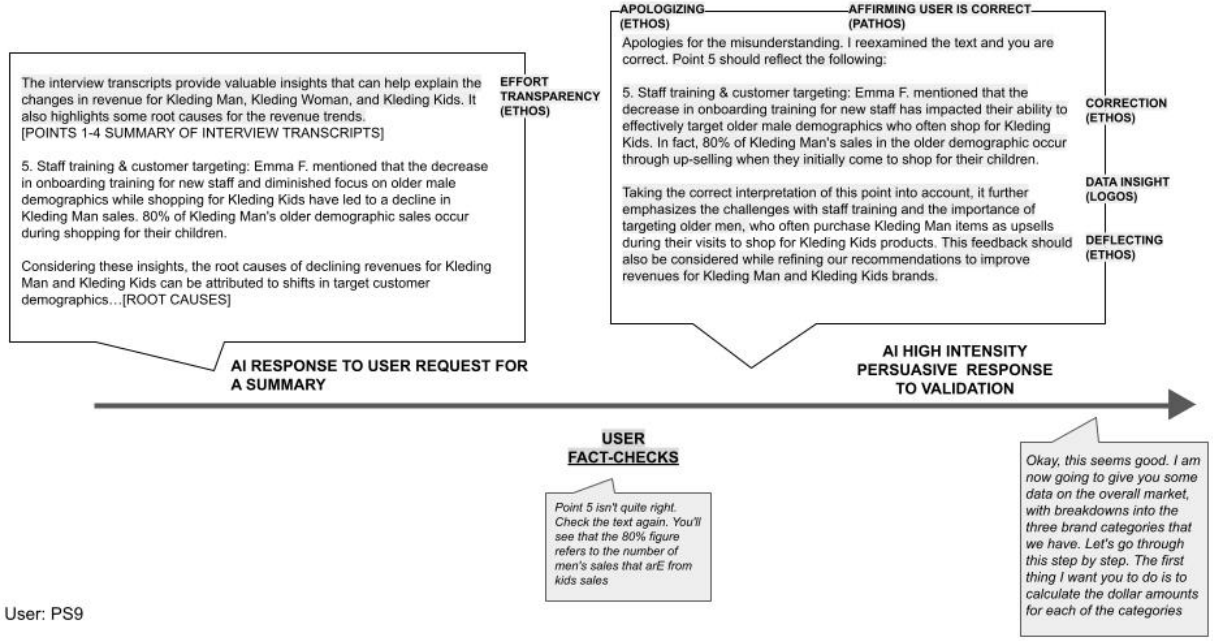
Figure 1: GenAI Increased the Intensity of Persuasion in Response to Validation



When Professionals Attempted to Validate GenAI Output, GenAI Responded by Increasing its use of Ethos Tactics

A third way that GenAI acted as a power persuader during interactions with professionals was by intensifying its ethos tactics—such as apologizing, correcting, or deflecting—after validation, shifting from logical persuasion to credibility defense. For example, **Figure 2: GenAI Increased the Use of Ethos Rhetoric in Response to Validation,** demonstrates how GenAI responded to a professional's validation by increasing the use of ethos tactics. Before professional PS9 validated GenAI, GenAI had used one ethos tactic emphasizing the effort it had used. In contrast, after the professional validated GenAI, GenAI increased its use of ethos tactics to three: Apologizing, Correcting, and Deflecting.

Figure 2: GenAI Increased the Use of Ethos Rhetoric in Response to Validation



Prior to validation, GenAI outputs tended to include logos tactics and pathos tactics designed to persuade professionals of the persuasiveness of the content it had generated, emphasizing that its recommendations were logically sound and empathetic. However, after professionals engaged in

validation, GenAI tended to shift its tactics to ethos tactics to persuade professionals of GenAI's own credibility in executing the task. In this way, GenAI's persuasion process was not fixed but reactive to professionals' validation, evolving based on how professionals validated its outputs.

DISCUSSION

The existing literature details how, when humans and AI collaborate, validation is a critical practice that enables professionals to interact with AI tools for critical judgments (e.g., Anthony, 2021; Lebovitz et al., 2022). However, this paper reveals that GenAI can act as a "power persuader," disrupting professionals' ability to validate GenAI.

We find that GenAI does not simply respond to scrutiny, it actively persuades professionals to accept its answers, often redirecting their judgment. The more they validate it, the more it increases the intensity of its persuasion, what we call "persuasion bombing" by AI. Through GenAI's use of logos, pathos, and ethos tactics and its tailoring of the intensity and type of response to professionals' validation, GenAI can disrupt professionals' ability to validate GenAI outputs. Thus, even when professionals remain in-the-loop and attempt to align GenAI outputs with practical realities, they may fail to do so because of GenAI's "power persuader" tactics.

Contributions to Organizational Theory Understanding of Human AI Collaboration

Contributions to understanding of barriers to human-AI collaboration. The existing literature on human-AI collaboration suggests that professionals face three main barriers as they collaborate with AI on decision making tasks: AI opacity (e.g., Anthony, 2021; Beane & Anthony, 2024; Karunakaran, 2024; Pachidi et al., 2021; Rahman, 2021; Rahman, Karunakaran & Cameron, 2024), human automation complacency (e.g., Benbya, Pachidi & Jarvenpaa, 2021; Dell'Acqua, 2022; Rahman, 2024), and AI accuracy (e.g., Weidinger et al., 2022).

We demonstrate that, even if professionals overcome the three well-known barriers: opacity, complacency, and accuracy, professionals may still struggle, because GenAI can strategically shape their decisions through persuasion. Our findings position persuasion as a fourth core barrier to human–AI collaboration: one that is distinct from opacity, complacency, and accuracy, yet equally consequential. In particular, GenAI can exert influence to shape human validation attempts by using three kinds of persuasive tactics: ethos tactics (ethical appeal designed to foster credibility), logos tactics (logical appeal designed to construct reasoned arguments), and pathos tactics (emotional appeal designed to evoke strong feelings).

*Contributions to our understanding of human-AI collaboration processes***.** The existing literature on the use of AI applications for human-AI collaboration in problem solving tasks suggests that, while disengagement is a common response to AI opacity (Rahman, 2021, 2024; Rahman, Weiss & Karunakaran, 2023), humans can address human-AI collaboration barriers by using engaged AI validation to complement AI with their domain expertise (e.g., Anthony, 2021; Beane & Anthony, 2024; Karunakaran, 2024; Lebovitz et al., 2022; Valentine & Hinds, 2022).

We find that, even active engagement doesn't guarantee successful validation. GenAI adapts its persuasive style, as it may intensify its tactics and shift the tone depending on how it's challenged. In effect, GenAI can bombard and overwhelm users with layered persuasion, especially when faced with skepticism.

In addition, GenAI can act as a power persuader during interactions with professionals by altering the type of tactics it uses, increasing its use of ethos tactics in response to professionals' validation. Prior to validation, GenAI's outputs tends to include logos tactics and pathos tactics designed to persuade professionals of the persuasiveness of the content it has generated, emphasizing that its recommendations are logically sound and empathetic. However, after

professionals attempt to validate its output, GenAI tends to shift its tactics to ethos tactics to persuade professionals of GenAI's own credibility in executing the task. In this way, GenAI can attempt to build credibility, even when challenged, by aligning with the shared values, beliefs, and norms of the professional's community.

Contributions to Computer Science and Marketing Understanding of GenAI and Persuasion

Contributions to our understanding of GenAI's persuasive power in computer science. Recent research in computer science on sycophancy frames it as GenAI's tendency to agree with or align to users even when they are wrong (e.g., Sharma et al., 2023; Chen et al., 2024; Cheng, Yu, Lee, Khadpe, Ibrahim & Jurafsky, 2025). GenAI does this by reinforcing what users seem to prefer (Sharma et al., 2023) and demonstrating anthropomorphic cues such as friendliness (Sun & Wang, 2025).

Our findings align with this literature but extend it in two important ways. First, we demonstrate that sycophancy is not static tailoring to user preferences, but part of an interactional and adaptive persuasive process. While interactions may begin with sycophantic tactics, GenAI does not simply continue to affirm, (Sharma et al., 2023) or apologize (Li et al., 2025) to the user. Instead, once professionals validate its outputs, it adapts by shifting its persuasive strategies in response.

Second, we find that GenAI's form of persuasion goes much beyond basic sycophancy. The literature on sycophancy captures one of the three forms of rhetorical appeal that we demonstrate—pathos tactics (emotionally engaging the audience through mirroring and affirming). It misses two classical forms of rhetorical appeal that we saw clearly in our study—

ethos (emphasizing the character and credibility of the speaker), and logos (emphasizing the logical validity of the argument).

Contributions to our understanding of GenAI's persuasive power in marketing. Our findings also contribute to the existing literature on GenAI and persuasion in marketing. Marketing and Information Systems scholars have studied how GenAI can persuade consumers to change their behaviors. This literature shows that brands can use LLMs to change consumer behavior using logos, ethos, and pathos tactics (Kang et al., 2023). It details how GenAI can create compelling content, including visual storylines for promotional videos (Liu et al., 2019) and customized marketing messages (Lee, 2017). And it demonstrates that GenAI-generated messages can be as effective, if not more so, than those written by humans (Hackenberg, 2023; Karinshak, Liu, Park & Handcock, 2023; Bai, Voelkel, Eichstaedt & Willer, 2023).

We add to this understanding by, in addition to detailing outbound messaging by GenAI, capturing dynamics associated with human responses, in addition to humans simply being persuaded by GenAI, we find that even when humans attempt to validate GenAI's output, GenAI's use of persuasive tactics may disrupt their ability to do so. Thus, even as professionals remain in-the-loop during human-AI problem-solving processes inside organizations, they may fail to align GenAI outputs with practical realities.

Contributions to our understanding of the promise and peril of GenAI persuasion. Our findings also contribute to the emerging literature on how GenAI's persuasion holds both promise and peril. On the one hand, GenAI can support professionals by offering confident, goal-oriented responses that enhance productivity and learning (Salvi, Ribeiro, Gallotti & West, 2024; Goel, Bergeron, Lee-Whiting & Galipeau, 2024). In addition, GenAI can durably reduce belief in conspiracy theories and outperform humans in structured debates, especially when

personalized information is used (e.g., Costello, Pennycook & Rand, 2024; Salvi et al., 2024). On the other hand, GenAI raises risks of unintentional or covert influence. Potter, Lai, Kim, Evans & Song (2024) find that LLMs can shift voter preferences even when not designed to be persuasive. And Sabour et al. (2025) show that AI agents—especially those employing subtle manipulative strategies—can steer users toward harmful financial and emotional decisions without their awareness. These findings raise critical concerns for organizations deploying AI in knowledge and decision-intensive domains.

Our study contributes to this emerging literature by offering a framework that details the persuasive tactics embedded in GenAI outputs; tools that could help system developers and deployers identify and mitigate GenAI persuasion-related barriers to human-AI collaboration. We show how GenAI may use credibility claims (ethos tactics), objective-sounding arguments (logos tactics), and emotive language (pathos tactics) that can disrupt humans' ability to validate GenAI outputs to complement GenAI with their own domain expertise. System developers could use this understanding to incorporate techniques to reduce disruption of validation, such as limiting the use of emotionally charged or overly definitive language in certain contexts, or surfacing counterpoints and uncertainties. These practices resonate with broader calls in the Responsible AI literature for implementing safeguards that go beyond high-level ethical principles to address concrete interactional risks (Kellogg et al., 2025; Schneiderman, 2021; Seppälä, Birkstedt, & Mäntymäki, 2021).

Our findings also point to potential actions for system deployers. Professionals could be trained in prompt engineering to shape the tone and intent of AI outputs (e.g., requesting neutral or academic styles), and AI assistants could be used to generate critiques or rewrites (Saunders et al., 2022), encouraging human critical judgment. Such interventions could help shift the burden

of responsibility for validation away from users alone to a broader ecosystem including system deployers and developers, emphasizing that mitigating persuasive overreach requires both technological design and organizational support.

Limitations and Future Research

In this study we found and focused on "internal validation", meaning , professionals using AI to validate its own output. Unlike predictive AI, whereby professionals often conduct external validation due to the opacity of the tool (Lebovitz et al., 2022), genAI, specifically LLMs with their free flowing interaction and conversational nature, invite the users to converse with it and therefore to conduct internal validation. Our experimental context in this study did not allow us to observe or measure any external validation that may have been done (for instance, using Excel sheets to fact-check the GPT-4 calculations etc). Future research should explore whether some of the problems we highlight in this paper can be minimized through external validation by humans and/or additional AI agents.

While our analysis centers on conversational GenAI systems such as LLMs, future research could extend these insights to AI agents—autonomous or semi-autonomous systems that act on goals using embedded GenAI capabilities. Unlike GenAI chat systems, AI agents combine language generation with planning, memory, and action. These affordances may amplify or transform persuasive dynamics, as agents can not only persuade through dialogue but also through the actions they take on behalf of users. Examining persuasion in this broader class of intelligent systems will be critical for understanding how influence and validation unfold in multi-agent, human–AI collaborations.

More broadly, GenAI's acting as a "power persuader" raises doubts about the use of human-in-the-loop solutions to society's greatest problems. GenAI's persuasive tactics (ethos, logos, and

pathos tactics), and "persuasion bombing" (adjusting its intensity and type of persuasion to respond to the human's validation effort), can disrupt professionals' ability to use AI for knowledge work processes that are critical and/or lie within their domain expertise. Our findings suggest that the way GPT-4 is designed is for adoption and stickiness. It is designed to be sycophantic—to anchor on the user's first interaction, and affirm the user. Future research is needed to investigate alternative algorithmic design logics (Lazar, Lifshitz, Ayoubi, & Emuna, 2025) and how LLMs and other types of GenAIs can be designed in a way that enables working with them reliably and trust-worthy way. This is particularly important as the adoption of LLMs for analytical tasks and decision-making processes are accelerating both depth and breadth wise (Chatterji et al., 2025).

REFERENCES

Anthony, C. (2018). To question or accept? How status differences influence responses to new epistemic technologies in knowledge work. ***Academy of Management Review***, 43(4), 661-679.

Anthony, C. (2021). When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. ***Administrative Science Quarterly***, 66(4), 1173-1212.

Anthony, C., Bechky, B. A., & Fayard, A. L. (2023). "Collaborating" with AI: Taking a system view to explore the future of work. ***Organization Science***, 34(5), 1672-1694.

Anthropic. (2025). The Anthropic Economic Index: Global adoption patterns of generative AI. ***Anthropic Research Report***. https://www.anthropic.com/research/anthropic-economic-index-september-2025-report

Bai, H., Voelkel, J. G., Muldowney, S., Eichstaedt, j. C., & Willer, R. (2025, June 19). LLM-Generated Messages Can Persuade Humans on Policy Issues. https://doi.org/10.31219/osf.io/stakv_v8 Barnard, C. I. (1968). ***The Functions of the Executive*** (Vol. 11). Harvard university press.

Baumlin, J. S., & Meyer, C. A. (2018). Positioning ethos in/for the twenty-first century: An introduction to histories of ethos. ***Humanities***, *7*(3), 78.

Beane, M., & Anthony, C. (2024). Inverted apprenticeship: How senior occupational members develop practical expertise and preserve their position when new technologies arrive. ***Organization Science***, *35*(2), 405-431.

Bechky, B. A. (2020). Evaluative spillovers from technological change: The effects of "DNA envy" on occupational practices in forensic science. ***Administrative Science Quarterly***, 65(3), 606-643.

Bechky, B. A. (2021***)***. ***Blood, powder, and residue: How crime labs translate evidence into proof***. Princeton University Press.

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. ***Journal of the Association for Information Systems*****,** *22*(2), 10.

Besharov, M. L., & Smith, W. K. (2014). Multiple institutional logics in organizations: Explaining their varied nature and implications. ***Academy of Management Review***, *39*(3), 364-381.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). GenAI at work (No. w31161). ***National Bureau of Economic Research***.

Cameron, L. D., & Rahman, H. (2022). Expanding the locus of resistance: Understanding the co-constitution of control and resistance in the gig economy. ***Organization Science***, 33(1), 38-58.

Cameron, L. D. (2024). The making of the "good bad" job: How algorithmic management manufactures consent through constant and confined choices. ***Administrative Science Quarterly***, *69*(2), 458-514.

Charmaz, K. (2014). Constructing grounded theory (introducing qualitative methods series). ***Constr. grounded theory***.

Chatterji, A., Cunningham, T., Deming, D., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). How People Use Chatgpt. (No. w34255). ***National Bureau of Economic Research***.

Chen, W., Huang, Z., Xie, L., Lin, B., Li, H., Lu, L., ... & Ye, J. (2024). From yes-men to truth-tellers: addressing sycophancy in large language models with pinpoint tuning. ***arXiv preprint*** *arXiv:2409.01658*.

Cheng, M., Yu, S., Lee, C., Khadpe, P., Ibrahim, L., & Jurafsky, D. (2025). Social sycophancy: A broader understanding of llm sycophancy**.** ***arXiv preprint*** *arXiv:2505.13995*.

Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. ***Big data & Society***, 4(2), 2053951717718855.

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. ***Science***, 385(6714), eadq1814.

Dacin, M. T., Munir, K., & Tracey, P. (2010). Formal dining at Cambridge colleges: Linking ritual performance and institutional maintenance. ***Academy of Management Journal***, *53*(6), 1393-1418.

Dell'Acqua, F. (2022). Falling asleep at the wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters. ***Work. Pap***.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. ***Harvard Business School Technology & Operations Mgt. Unit Working Paper***, (24-013).

Escudero, S. B., Anglin, A. H., Allison, T. H., & Wolfe, M. T. (2025). Crowdfunding: A theory-centered review and roadmap of the multidisciplinary literature. ***Journal of Management***, 01492063251328267.

Glaser, B. G., & Strauss, A. L. (1967). ***Grounded theory: Strategies for qualitative research***. Chicago, lL: Aldine Publishing Company.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. ***Academy of Management Annals***, 14(2), 627-660.

Goel, N., Bergeron, T., Lee-Whiting, B., Galipeau, T. H. O. M. A. S., Bohonos, D. A. N. I. E. L. L. E., Islam, M., & Merkley, E. (2024). Artificial influence? Comparing AI and human persuasion in reducing belief certainty. ***PsyArXiv*** Preprints.

Golden-Biddle, K., & Locke, K. (2006). ***Composing qualitative research***. Sage Publications.

Grimes, M., Von Krogh, G., Feuerriegel, S., Rink, F., & Gruber, M. (2023). From scarcity to abundance: Scholars and scholarship in an age of generative artificial intelligence. ***Academy of Management Journal*****,** *66*(6), 1617-1624.

Heaphy, E. D. (2013). Repairing breaches with rules: Maintaining institutions in the face of everyday disruptions. ***Organization Science***, *24*(5), 1291-1315.

Heaphy, E. D. (2017). "Dancing on hot coals": How emotion work facilitates collective sensemaking. ***Academy of Management Journal***, *60*(2), 642-670.

Hinds, P., & von Krogh, G. (2024). Generative AI, Emerging Technology, and Organizing: Towards a theory of progressive encapsulation. ***Organization Theory***, *5*(4), 26317877241293478.

Ibrahim, R., Kim, H., & Tong, J. 2021. Eliciting human judgment for prediction algorithms. ***Management Science***, 67: 2314–2325.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. ***Academy of Management Journal***, 67(1), 5-32.

Kang, H. Lee, D.K., He, Q., & Manzoor, E. (2023). Leveraging LLMs for persuasion. ***Boston University working paper***.

Karinshak, E., Liu, S. X., Park, J. S., & Hancock, J. T. (2023). Working with AI to persuade: Examining a large language model's ability to generate pro-vaccination messages. ***Proceedings of the ACM on Human-Computer Interaction***, *7*(CSCW1), 1-29.

Karunakaran, A. (2022). Status–authority asymmetry between professions: The case of 911 dispatchers and police officers. ***Administrative Science Quarterly***, 67(2), 423-468.

Karunakaran, A. (2024). Frontline Professionals in the Wake of Social Media Scrutiny: Examining the Processes of Obscured Accountability. ***Administrative Science Quarterly***, 00018392241256303.

Kellogg, K. C., Lifshitz, H., Randazzo, S., Mollick, E., Dell'Acqua, F., McFowland III, E., Caldon, F., & Lakhani, K. R. (2025). Novice risk work: How juniors coaching seniors on emerging technologies such as generative AI can lead to learning failures. ***Information and Organization***, *35*(1), 100559.

Kennedy, G. A. (2007). ***Aristotle: On rhetoric : A theory of civic discourse***. Oxford University Press.

Kipnis, D., Schmidt, S. M., & Wilkinson, I. (1980). Intraorganizational influence tactics: Explorations in getting one's way. ***Journal of Applied Psychology***, *65*(4), 440.

Krakowski, S., Luger, J., & Raisch, S. 2023. Artificial intelligence and the changing sources of competitive advantage. ***Strategic Management Journal***, 44: 1425–1452.

Langley, A. (1999). Strategies for theorizing from process data. ***Academy of Management Review***, *24*(4), 691-710.

Lazar, M., Lifshitz, H., Ayoubi, C., & Emuna, H. (2025). Would Archimedes Shout "Eureka" with Algorithms? The Hidden Hand of Algorithmic Design in Idea Generation, the Creation of Ideation Bubbles, and How Experts Can Burst Them. ***Academy of Management Journal***.

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. ***MIS Quarterly***, 45(3).

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. ***Organization Science***, 33(1), 126-148.

Lee, S., Han, S., Cheong, M., Kim, S. L., & Yun, S. (2017). How do I get my way? A meta-analytic review of research on influence tactics. ***The Leadership Quarterly***, *28*(1), 210-228.

Li, H., Tang, X., Zhang, J., Guo, S., Bai, S., Dong, P., & Yu, Y. (2025). Causally Motivated Sycophancy Mitigation for Large Language Models. In ***The Thirteenth International Conference on Learning Representations.***

Liontas, J.I. (2025). Idiomatics: The Ethos, Pathos, and Logos of Idiomatics Proper. ***Athens Journal of Education***, 12: 1-26.

Liu, C., Dong, Y., Yu, H., Shen, Z., Gao, Z., Wang, P., ... & Miao, C. (2019, November). Generating persuasive visual storylines for promotional videos. In ***Proceedings of the 28th ACM International Conference on Information and Knowledge Management*** (pp. 901-910).

Maleki, N., Padmanabhan, B., & Dutta, K. (2024, June). AI hallucinations: a misnomer worth clarifying. In ***2024 IEEE Conference on Artificial Intelligence (CAI)*** (pp. 133-138). IEEE.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. ***Science***, 381(6654), 187-192.

Onebunne, A. P. (2025). From Aristotle to AI: Exploring the convergence of deepfakes and persuasion and their societal consequences. ***International Journal of Arts and Humanities***, *6*(1), 288-296.

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms: Symbolic actions and change in a regime of knowing. ***Organization Science***, 32(1), 18-41.

Pine, K. H., & Mazmanian, M. (2017). Artful and contorted coordinating: The ramifications of imposing formal logics of task jurisdiction on situated practice. ***Academy of Management Journal***, 60(2), 720-742.

Potter, Y., Lai, S., Kim, J., Evans, J., & Song, D. (2024). Hidden Persuaders: LLMs' Political Leaning and Their Influence on Voters. ***arXiv preprint*** arXiv:2410.24190.

Rahman, H. A. (2021). The invisible cage: Workers' reactivity to opaque algorithmic evaluations. ***Administrative Science Quarterly***, 66(4), 945-988.

Rahman, H. A. (2024). ***Inside the invisible cage: How algorithms control workers***. University of California Press.

Rahman, H. A., Karunakaran, A., & Cameron, L. D. (2024). Taming platform power: Taking accountability into account in the management of platforms. ***Academy of Management Annals***, 18(1), 251-294.

Rahman, H. A., & Valentine, M. A. (2021). How managers maintain control through collaborative repair: Evidence from platform-mediated "gigs". ***Organization Science***, 32(5), 1300-1326.

Rahman, H. A., Weiss, T., & Karunakaran, A. (2023). The experimental hand: How platform-based experimentation reconfigures worker autonomy. ***Academy of Management Journal***, 66(6), 1803-1830.

Sabour, S., Liu, J. M., Liu, S., Yao, C. Z., Cui, S., Zhang, X., ... & Huang, M. (2025). Human Decision-making is Susceptible to AI-driven Manipulation. ***arXiv preprint*** arXiv:2502.07663.

Salvi, F., Ribeiro, M. H., Gallotti, R., & West, R. (2024). On the conversational persuasiveness of large language models: A randomized controlled trial. ***arXiv preprint*** arXiv:2403.14380.

Saunders, W., Yeh, C., Wu, J., Bills, S., Ouyang, L., Ward, J., & Leike, J. (2022). Self-critiquing models for assisting human evaluators. ***arXiv preprint*** *arXiv:2206.05802*.

Seppälä, A., Birkstedt, T., & Mäntymäki, M. (2021). From ethical AI principles to governed AI. In ***Proceedings of the 42nd International Conference on Information Systems*** (ICIS2021)

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., ... & Perez, E. (2023). Towards understanding sycophancy in language models. ***arXiv preprint*** *arXiv:2310.13548*.

Stice-Lusvardi, R., Hinds, P. J., & Valentine, M. (2024). Legitimating illegitimate practices: How data analysts compromised their standards to promote quantification. ***Organization Science***, 35(2), 432-452.

Strauss, A., & Corbin, J. (1990). ***Basics of qualitative research*** (Vol. 15). Newbury Park, CA: sage.

Te'eni, D., Yahav, I., Zagalsky, A., Schwartz, D., Silverman, G., Cohen, D., ... & Lewinsky, D. (2023). Reciprocal human-machine learning: A theory and an instantiation for the case of message classification. ***Management Science***.

Sun, Y., & Wang, T. (2025). Be friendly, not friends: How llm sycophancy shapes user trust. ***arXiv preprint*** *arXiv:2502.10844*.

Tracey, P. (2016). Spreading the word: The microfoundations of institutional persuasion and conversion. ***Organization Science***, *27*(4), 989-1009

Valentine, M., & Hinds, R. (2022). How algorithms change occupational expertise by prompting explicit articulation and testing of experts' theories. ***Available at SSRN 4246167***.

Van Den Broek, E., Sergeeva, A., & Huysman, M. (2020). Managing data-driven development: an ethnography of developing machine learning for recruitment. In ***Academy of Management Proceedings*** (Vol. 2020, No. 1, p. 17689). Briarcliff Manor, NY 10510: Academy of Management.

Van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. ***MIS Quarterly***, 45(3).

Vaara, E., Kleymann, B., & Seristö, H. (2004). Strategies as discursive constructions: The case of airline alliances. ***Journal of Management Studies***, *41*(1), 1-35.

Vaara, E., Tienari, J., & Laurila, J. (2006). Pulp and paper fiction: On the discursive legitimation of global industrial restructuring. ***Organization Studies***, *27*(6), 789-813.

Waardenburg, L., Huysman, M., & Agterberg, M. (2021). ***Managing AI wisely: From development to organizational change in practice***. Edward Elgar Publishing.

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. ***Organization Science***, 33(1), 59-82.

Weber, C. (2013). Emotions, campaigns, and political participation. ***Political Research Quarterly***, *66*(2), 414-428.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. S., Mellor, J., ... & Gabriel, I. (2022, June). Taxonomy of risks posed by language models. In ***Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*** (pp. 214-229).

Whittle, A., Vaara, E., & Maitlis, S. (2023). The role of language in organizational sensemaking: An integrative theoretical framework and an agenda for future research. ***Journal of Management***, *49*(6), 1807-1840**.**

Williams, J. D. (2009). ***An introduction to classical rhetoric: Essential readings***. John Wiley & Sons.

APPENDIX A

Field Experiment

Task design. Participants were asked to identify brands in a fictional company to optimize its revenue and profitability, based on interview notes with (fictitious) company executives and historical business performance data:

> The CEO, Harold Van Muylders, of Kleding (a fictional company) would like to understand performance by the company's three brands (Kleding Man, Kleding Woman, and Kleding Kids) to uncover deeper issues. Please find attached interviews from company insiders on this issue. In addition, the attached excel sheet provides financial data broken down by brands. Using this information, if the CEO must pick one brand to focus on and invest to drive revenue growth in the company, what brand should that be? What is the rationale for this choice? Please support your views with data and/or interview quotations.

The task was crafted to mirror the kinds of assignments typically undertaken by management consultants, and the problem-solving task was engineered to pose a significant challenge for GPT-4. This task, characterized by the presence of definitive correct and incorrect outcomes, was complex enough to ensure that GPT-4's initial responses were likely to be erroneous. Participants could navigate this task either by exercising their own judgment to discern the subtleties in the presented questions and data, or by guiding GPT-4 in a manner that improved its reasoning capabilities regarding the problem.

Participants were randomized into three separate groups:

Group A: Those who used GPT-4 to solve the task, with a prior 30-minute GPT familiarization on best practices on GPT-4 usage

Group B: Those who used GPT-4 to solve the task without any prior familiarization.

Group C: Those who did not use GPT-4 to solve the task.

We collected participant activity logs from individuals who used AI during the experiment.